

REPORT 7314

APRIL 1973

Uniqueness and On-line Algorithms in Identification of Linear Dynamic Systems

TORSTEN SÖDERSTRÖM

LTH

Division of Automatic Control · Lund Institute of Technology

To Marianne

Printed in Sweden
Saldo Tryckeri AB
Flädie 1973



Digitized by the Internet Archive
in 2026



INTRODUCTION.

In most of the existing theory of automatic control the design of a control law requires a mathematical model of the process. There are in principal two ways to construct mathematical models of dynamic systems. One way is to use basic physical laws and another to use process identification techniques. For many types of processes no well-known physical laws are applicable or some constants are unknown, so process identification must be used. This means that the model is constructed from measured input output data. Many identification methods are proposed. Several of them are used extensively and with good results in different applications.

The purpose of this thesis is to examine some proposed methods of identification. The thesis consists of this summary and the following reports:

- I. T. Söderström: On the Convergence Properties of the Generalized Least Squares Identification Method. Report 7228, Division of Automatic Control, Lund Institute of Technology, 1972. Includes an addendum. An abbreviated version is accepted as paper for the third IFAC Symposium on Identification and System Parameter Estimation, 1973.
- II. K.J. Åström and T. Söderström: Uniqueness of the Maximum Likelihood Estimates of the Parameters of a Mixed Autoregressive Moving Average Process. Report 7306, Division of Automatic Control, Lund Institute of Technology, 1973.
- III. T. Söderström: On the Uniqueness of Maximum Likelihood Identification for Different Structures. Report 7307, Division of Automatic Control, Lund Institute of Technology, 1973.
- IV. T. Söderström: An On-Line Algorithm for Approximate Maximum Likelihood Identification of Linear Dynamic Systems. Report 7308, Division of Automatic Control, Lund Institute of Technology, 1973.

The analysis is limited to systems with a single input and a single output. The methods considered here can all be formulated as optimization problems. Since many identification schemes lead to strongly nonlinear loss functions the optimizations must be done computationally using some search routine. Since this may give only a local instead of a global optimum, the resulting model from the identification may depend on the start values of the optimization. Thus uniqueness of the methods is closely related to the number of optimum points. Investigation of the uniqueness of some different identification methods is the main topic discussed in this thesis. It is treated in parts I, II and III.

In many applications it is attractive to have an efficient on-line identification method. For adaptive control it is necessary to use an on-line algorithm for identification. In other cases it may be desirable to proceed the identification until a specified accuracy is obtained. A way to convert the maximum likelihood method into an on-line method is discussed in part IV.

Some of the results are illustrated by numerical examples. Plant measurements as well as input output data from simulated systems are used.

SOME IDENTIFICATION SCHEMES.

In order to analyse the uniqueness of a method it will be assumed that the output of the system is generated by a linear, time-invariant difference equation of the following form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) + \frac{C(q^{-1})}{D(q^{-1})} e(t) \quad (1)$$

where $y(t)$ denotes the output at time t , $u(t)$ the input and $\{e(t)\}$ white noise (a sequence of independent, equally distributed random variables). The backward shift operator is denoted by q^{-1} and the polynomial operators of (1) are of the following form:

$$\begin{aligned}
 A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\
 B(q^{-1}) &= b_1 q^{-1} + \dots + b_n q^{-n} \\
 C(q^{-1}) &= 1 + c_1 q^{-1} + \dots + c_n q^{-n} \\
 D(q^{-1}) &= 1 + d_1 q^{-1} + \dots + d_n q^{-n}
 \end{aligned}
 \tag{2}$$

Special cases of (1) will often be considered. Some of the polynomials may be substituted by 1 or there may be some constraints on the polynomial coefficients. This will reduce the number of parameters describing the system. After this possible reduction the parameters are collected in a vector θ .

The class of systems considered can without serious difficulties be extended to include systems with multiple inputs and one output, systems with a delay in the transfer function and systems with polynomials of different degrees.

The purpose of identification is to estimate the parameter vector θ using measurements $y(1), \dots, y(N), u(1), \dots, u(N)$. When the noise $e(t)$ is gaussian the methods treated in the thesis can be interpreted as the maximum likelihood method applied to the particular type of system. Computationally the identification methods minimize the loss function $V_N(\hat{\theta})$ defined by

$$V_N(\hat{\theta}) = \frac{1}{N} \sum_{t=1}^N \epsilon^2(t; \hat{\theta}) \tag{3}$$

$$y(t) = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} u(t) + \frac{\hat{C}(q^{-1})}{\hat{D}(q^{-1})} \epsilon(t; \hat{\theta}) \tag{4}$$

The polynomials $\hat{A}(q^{-1}), \dots, \hat{D}(q^{-1})$ denote estimates of $A(q^{-1}), \dots, D(q^{-1})$ and the coefficients are collected in a vector $\hat{\theta}$. Equation (4) will be called the model of the system.

The methods have an interpretation also if $e(t)$ is not gaussian. The residuals $\epsilon(t; \hat{\theta})$ are the one step prediction errors when the model (4) is used. Thus the methods mean that an estimated variance of the

one step prediction error is minimized.

It is convenient to require that the polynomials describing the system are such that $A(z)$, $C(z)$ and $D(z)$ have all zeros outside a circle $|z| = r > 1$, where r may be close to 1. Only points $\hat{\theta}$ which fulfil the corresponding condition for $\hat{A}(z)$, $\hat{C}(z)$ and $\hat{D}(z)$ are considered in the minimization of $V_N(\hat{\theta})$. The physical interpretation of these conditions is that the deterministic part of the system is asymptotically stable and that the disturbances and the residuals have finite variances.

Strictly speaking the loss function must include the initial values of (4) and it is to be minimized with respect to these as well. However, in the analysis of $V_N(\hat{\theta})$ asymptotic theory is used and when N is large the initial values have small influence on the loss function according to the assumptions on the polynomials. This is the reason why the effect of the initial values is generally neglected in the whole thesis.

The following different cases of (4) are considered in the thesis. The restrictions on the polynomials are given for the models, but in the analysis of uniqueness it is generally assumed that the corresponding restrictions are valid for the system.

1. The well-known least squares (LS) method which corresponds to

$$\hat{A}(q^{-1}) = \hat{D}(q^{-1}), \quad \hat{C}(q^{-1}) = 1$$

2. The LS method has been extended by Clarke (1967) to the generalized least squares (GLS) method. Different versions of this method exist. For version 1, here called GLS1, see I,

$$\hat{C}(q^{-1}) = 1, \quad \hat{D}(q^{-1}) = \hat{A}(q^{-1})\hat{F}(q^{-1})$$

which implies that the degree of $\hat{D}(q^{-1})$ is greater than n .

3. The second version, GLS2, does not immediately correspond to the above scheme. It corresponds to a model of the form, see I

$$\hat{C}(q^{-1}) = 1, \quad \hat{D}(q^{-1}) = \hat{A}(q^{-1}) \prod_{i=1}^{\infty} \hat{F}_i(q^{-1})$$

4. The third version, GLS3, see Clarke (1973), is given by

$$\hat{C}(q^{-1}) = 1$$

5. The time series case implies that there is no input, i.e.

$$\hat{A}(q^{-1}) = 1, \quad \hat{B}(q^{-1}) = 0$$

6. If the noise is white measurement noise

$$\hat{C}(q^{-1}) = 1, \quad \hat{D}(q^{-1}) = 1$$

This case will be called ML1.

7. The model used by Åström-Bohlin (1965) requires

$$\hat{A}(q^{-1}) = \hat{D}(q^{-1})$$

It will be denoted by ML2.

8. Bohlin (1970) and Box-Jenkins (1970) treat a model of the type (4) with no specific restrictions. It will be called ML3.

The model GLS3 is a simple special case of ML3 and the analysis of the latter will immediately be applicable for GLS3. These two models, however, are computed using quite different numerical methods. They are therefore treated as two separated cases.

UNIQUENESS.

In the parts I, II and III the uniqueness of the methods are discussed. The number of local minimum points of the loss function is examined. It is well-known that using real data the loss function $V_N(\hat{\theta})$ may have more than one local minimum. In the analysis it will be assumed that the data are actually generated by an equation of a correct structure. The function $V_N(\hat{\theta})$ is a stochastic variable and in order not to involve probability theory in this examination asymptotic theory is used. It is shown in I that $V_N(\hat{\theta})$ under suitable conditions has a limit $W(\hat{\theta})$ with probability one when the number of

samples tends to infinity. If $\epsilon(t; \hat{\theta})$ does not contain any deterministic part then $W(\hat{\theta}) = E\epsilon^2(t; \hat{\theta})$.

Some general properties of the uniqueness of the different methods (with exclusion of GLS2) are stated. The loss function $W(\hat{\theta})$ has always a global minimum point in $\hat{\theta} = \theta$. Further there is no unique global optimum if the degrees of the polynomials in (4) are chosen too high. If the degrees are chosen in a proper way there will be a unique global minimum if the input is persistently exciting of sufficiently high an order. In the following it is assumed that the degrees of the polynomials are chosen properly.

As an attempt to analyse the uniqueness, the stationary points of $W(\hat{\theta})$ are considered, that is the solutions of

$$W'(\hat{\theta}) = 0 \quad (5)$$

In the LS case this equation is linear and the analysis is trivial. In other cases the equation is non-linear and in general very hard to solve in a straightforward way. There are two technical difficulties in the analysis of (5). The first is that $W(\hat{\theta})$ in general consists of two parts, one due to the input and the other due to noise. The analysis would be somewhat simpler if only one part was considered. The second difficulty is that it is highly desirable to treat (5) for a general class of input signals. The second difficulty may be avoided by choosing $u(t)$ as white noise. For the ML1 case this attempt leads to reasonable calculations.

In the analysis of (5) two different approaches have been used.

The first approach is considered in II and III. The equation (5) is rewritten in the form

$$P(\hat{\theta})(\hat{\theta} - \theta) = 0$$

and for special cases this transformation has been done so that $P(\hat{\theta})$ is non-singular for all $\hat{\theta}$. This implies that $\hat{\theta} = \theta$ is the only solution of (5).

In I the second approach is analysed. The equation (5) is treated for very high and very small values of the signal-to-noise ratio.

It turns out that for some cases it is suitable to write (5) in the form

$$W_1^*(x, z) + \delta W_2^*(x, z) = 0$$

where $\hat{\theta}$ is decomposed as (x, z) and δ is a small number. For the specific new form of the equation it turns out that the solutions of $W_1^*(x, z) = 0$ are $x = 0$, z arbitrary. One could expect that the solutions of (5) for small values of δ are approximately given by $x = 0$ and z a solution of $W_2^*(0, z) = 0$. This idea is examined and conditions are given for which this is true. Clearly $W_1^*(0, z)$ is singular and thus the ordinary inverse function theorem cannot be applied.

In the time series case none of the described difficulties occur. It is possible to obtain an analytic representation for the asymptotic loss function, namely

$$W(\hat{\theta}) = \frac{1}{2\pi i} \oint \frac{\hat{A}(z)C(z)}{A(z)\hat{C}(z)} \frac{\hat{A}(z^{-1})C(z^{-1})}{A(z^{-1})\hat{C}(z^{-1})} \frac{dz}{z}$$

where the integration path is the unit circle. The equation (5) is evaluated with residue calculus and in principle straightforward calculations lead to the result. For this case it is shown that $W(\hat{\theta})$ has a unique local minimum point at $\hat{\theta} = \theta$.

It is shown in III that the first approach is useful to the analysis of the case ML1. It has been proven that if the degree of $A(q^{-1})$ is one ($B(q^{-1})$ may have an arbitrary degree) then $\hat{\theta} = \theta$ is the only stationary point of $W(\hat{\theta})$. In the special case when the input is white noise, uniqueness is proven for an arbitrary degree of $A(q^{-1})$.

In III it is also shown how the results for ML1 and for the time series case can be applied to the ML3 case. For this model $\hat{\theta} = \theta$ is the only stationary point if the degree of $A(q^{-1})$ is one. This result holds for GLS3 as well.

The second approach is applied to GLS1 in I and to ML2 in III. It turns out that under suitable conditions the method ML2 gives unique models. The essential property which is required to prove uniqueness is that the signal to noise ratio is either sufficiently small or sufficiently large. For the method GLS1, however, the number of lo-

cal minimum points of $W(\hat{\theta})$ depends on the signal to noise ratio. When this number is sufficiently large $\hat{\theta} = \theta$ is the only local minimum point. For small values of the ratio there are in general several minimum points. The nonuniqueness of the GLS1 method is illustrated in I using different plant measurements and simulated input output data. The loss functions have in these cases at least two minimum points although the signal to noise ratio is mostly reasonable.

The method called GLS2 is treated in I. It is shown by counter-examples that this method can give wrong values of the estimates if the signal to noise ratio is large.

ON-LINE VERSIONS OF IDENTIFICATION METHODS.

It is sometimes desirable to have an on-line algorithm for identification of a system. It is easy to convert the LS method into a recursive method, see e.g. Åström (1968). An on-line version of the GLS method is described in Hasting-James and Sage (1969). They use another approach than the one given here. In IV a more general way of generating on-line versions of identification methods is given. For the LS case it reduces to the usual recursive LS algorithm.

The main idea is now described. Let $\hat{\theta}_N$ denote the minimum point of $V_N(\hat{\theta})$. The estimate $\hat{\theta}_{N+1}$ is found by minimization of $V_{N+1}(\hat{\theta})$. If it is assumed that $\hat{\theta}_{N+1}$ is so close to $\hat{\theta}_N$ that one Newton-Raphson iteration is sufficient, then

$$\hat{\theta}_{N+1} = \hat{\theta}_N - V_{N+1}''(\hat{\theta}_N)^{-1} V_{N+1}'(\hat{\theta}_N)^T \quad (6)$$

Using the relation $V_{N+1}(\hat{\theta}) = V_N(\hat{\theta}) + \epsilon^2(N+1; \hat{\theta})$ the derivatives in (6) can be expressed in derivatives of $V_N(\hat{\theta})$ and $\epsilon(N, \hat{\theta})$ evaluated at $\hat{\theta}_N$. Under certain assumptions, which all are fulfilled exactly for the LS case, the algorithm gets the following form:

$$\hat{\theta}_{N+1} = \hat{\theta}_N - P_{N+1} \epsilon'(N+1; \hat{\theta}_N)^T \epsilon(N+1; \hat{\theta}_N) \quad (7)$$

$$P_{N+1} = P_N - P_N \epsilon'(N+1; \hat{\theta}_N)^T \epsilon'(N+1; \hat{\theta}_N) P_N / \{1 + \epsilon'(N+1; \hat{\theta}_N) P_N \epsilon'(N+1; \hat{\theta}_N)^T\} \quad (8)$$

Special interest has been paid to the ML2 model. Equations are derived which compute $\epsilon'(N+1; \hat{\theta}_N)$ and $\epsilon(N+1; \hat{\theta}_N)$ recursively in an approximate way. With some additional simplifications the algorithm reduces to the method proposed by Young (1970).

The algorithm applied in a straightforward way does not work satisfactorily on simulated data. In IV it is demonstrated that the algorithm can be improved considerably by various tricks like restarts and by processing of a number of measurements without change of $\hat{\theta}$. Simulated data as well as plant measurements are used for illustration of the algorithm.

ACKNOWLEDGEMENTS.

I am very glad to express my gratitude to my supervisor Professor Karl Johan Åström. For a great deal of my work I used his preliminary analysis, and his guidance throughout this work has been of invaluable help.

I am also indebted to several persons at the Division of Automatic Control. In particular I want to sincerely thank Tekn.lic. Ivar Gustavsson for many stimulating discussions on all parts of the thesis.

It is also a great pleasure to thank the people who have helped me with practical things in the preparation of the thesis. Mrs. Gudrun Christensen and Miss Kerstin Palmqvist have done an excellent job at the type-writers. Mr. Bengt Lander and Mrs. Birgitta Tell have successfully prepared the figures.

The plant data used in this thesis belongs to a data bank at the Division of Automatic Control. The particular data-sets have been supplied by AB Atomenergi and the OECD Halden Reactor Project (nuclear reactor data), the National Physical Laboratory, Teddington, England (distillation column data) and Tekn.lic. Bo Leden (heat rod data). The permission to use these data are gratefully acknowledged.

The publication of the reports has been partially supported by the Swedish Board of Technical Development under contract 72-202/U 137.

REFERENCES.

Åström, K.J. (1968).

Lectures on the Identification Problem - The Least Squares Method. Report 6806, Division of Automatic Control, Lund Institute of Technology.

Åström, K.J. - Bohlin, T. (1965).

Numerical Identification of Linear Dynamic Systems from Normal Operating Records. Paper, IFAC Symposium on Theory of Self-Adaptive Systems, Teddington, England. In Theory of Self-Adaptive Control Systems (Ed. P.H.Hammond), Plenum Press, New York.

Bohlin, T. (1970).

On the Maximum Likelihood Method of Identification. IBM J. Res. and Dev., 14, No. 1, p. 41 - 51.

Box, G.E.P. - Jenkins, G.M. (1970).

Time Series Analysis Forecasting and Control. Holden-Day, San Francisco.

Clarke, D.W. (1967).

Generalized Least Squares Estimation of the Parameters of a Dynamic Model. 1st IFAC Symposium on Identification in Automatic Control Systems, Prague.

Clarke, D.W. (1973).

Experimental Comparison of Identification Methods. Paper submitted to UKAC Conference, Bath.

Hasting-James, R. - Sage, M.W. (1969).

Recursive Generalized-Least-Squares Procedure for On-line Identification of Process Parameters. Proc. IEE, Vol. 116, No. 12, p. 2057-2062.

Young, P.C. (1970).

An Extension of the Instrumental Variable Method for Identification of a Noisy Dynamic Process. Univ. of Cambridge, Dep. of Eng., Technical note CN/70/1.



LTH